XX ENMC
ENCONTRO NACIONAL DE MODELAGEM COMPUTACIONAL
VIII ECTM
ENCONTRO DE CIÊNCIA E TECNOLOGIA DE MATERIAIS

16 a 19 de Outubro de 2017
Instituto Politécnico - Universidade do Estado de Rio de Janeiro
Nova Friburgo - RJ

Audiovisual Analytics Vocabulary and Ontology (AAVO): initial core and example expansion

Renato Fabbri - renato.fabbri@gmail.com
Maria Cristina Ferreira de Oliveira - cristina@icmc.usp.br
University of São Paulo, Institute of Mathematical and Computer Sciences - São Carlos, SP, Brazil

Abstract. *Visual Analytics might be defined as data mining though interactive visual interfaces. The field and has received noticeable consideration by researchers and the industry. The literature, however, is complex because it involves many fields of knowledge and is considerably recent. In this article we describe an initial organization of this knowledge as an OWL ontology and a SKOS vocabulary. This effort might be useful in many ways that include conceptual considerations, software implementations. Within the results and discussions, we expose a core and an example expansion of the conceptualization, and incorporate design issues that enhance the expressive power of the abstraction.*

Keywords: *OWL, SKOS, Semantic web, Visual analytics, Data visualization*

1. INTRODUCTION

Visual Analytics' usual definition is data mining (or the science of analytical reasoning facilitated) through interactive visual interfaces. From this definition, one can grasp that at least three fields are directly related to visual analytics: data mining, human-computer interaction, and data visualization. Each one of these fields are multidisciplinary and known to be considerably complex with multiple theories and vast literature.

This work is proposed as an organization of the knowledge related to visual analytics as a SKOS vocabulary and an OWL ontology. In other words, this document reports in an initial formalized conceptualization in terms of linked data recommended technologies. The uses this formalization can have includes:

- introduction of the Visual Analytics subject for non-specialists.
- Expressing a concise overview.
- Queries.
- Relating objects (e.g. data and techniques) inside the field or between the field and other domains.

- Making inference about the concepts and objects to which they are related.

In the present case, where the formalization is done as linked data, the conceptualization allows the queries, inferences, relations, etc. to be performed also by machines. Therefore, for example, a software might be able to relate a dataset or analysis methods to visualization techniques to aid a user of for automated reporting.

Section 2 uncover the methods. Section 3 is dedicated to stating and discussing the results. Final remarks and future work are in Section 4.

2. METHODS

The methods described here are standard of the semantic web to achieve formalized conceptualizations. The next sections address the subjects very briefly and the interested reader can visit the literature of the field and will surely find great items, including the W3C recommendations (W3C, 2010; Heath & Bizer, 2011).

2.1 The semantic web

The semantic web is constituted by data which are linked in the same way as web pages are: through HTTP and URLs. W3C recommendations are the main sources of protocols and best practices of the field. In practice, the terms 'linked data' and 'semantic web' are most often used interchangeably. A distinction might arise in some contexts where on need to refer to the linked data or the semantic web created by all or some portion of linked data, but, as a knowledge and technological field, the terms are equivalent. The main topics of this article are data visualization (or visual analytics) and semantic web. Accordingly, all the sections tackle the subject of the semantic web.

2.2 RDF

The semantic web is built using the Resource Description Framework (RDF). The RDF data model is based on making statements in the form of triples ("subject-predicate-object") and using Unique Resource Identifier for objects and concepts. It is also part of the framework to use URIs that are URLs whenever possible, to enable the data linkage. Accordingly, one can write:

```
<http://example.org/people/mary>
        <http://example.org/properties/name> "Mary Shastacian" .
<http://example.org/people/mary>
        <http://example.org/properties/age> "57" .
<http://example.org/people/mary>
        <http://example.org/properties/likes>
                <http://example.org/concepts/Reading> .
```

to express that there is person called Mary Shastacian which is 57 years old and likes reading. There are many formats to write/serialize RDF data. The example above is written in Turtle format and it will be used throughout this document.

In real settings, when everything is working as recommended, each of these URIs (that are URLs!) might be accessed through HTTP to access more triples referring to the URI accessed. From the triples above, one would be able to access triples describing each of the properties:

`example:properties/name`, `example:properties/age`, and `example:properties/likes`, and the concept `example:concepts/Reading`. In fact, the triples above would probably be available in the URI/URL: `example:people/mary`. One good working example[1] of this is DBPedia (Lehmann et al., 2015). The process of accessing a URI to find more triples is called *dereferencing the URI* (or simply dereferencing).

2.3 RDFS

The Resource Description Framework Schema (RDF Schema or simply RDFS) is a set of classes and properties for the RDF data model that allows basic descriptions of ontologies. It supports taxonomic relations (hypernymy, i.e. relations stating that a concept is more general than another), bindings of properties to objects and datatypes, and notes (label, comment, see also, is defined by).

2.4 SKOS

The Simple Knowledge Organization System (SKOS) is a data model for representing controlled vocabularies. SKOS is a W3C recommendation to facilitate publication and use of vocabularies and is built upon RDF and RDFS. It holds itself a vocabulary for concepts, notation, documentation, semantic and mapping relations, and collections.

2.5 OWL

The Web Ontology Language (OWL) is a language for publishing ontologies on the web. While RDFS holds basic relations necessary even for very rudimentary organization knowledge and data, OWL is complex and allows one to formalize elaborate conceptualizations. Using OWL, an ontology might have properties that are required to satisfy a number or axioms, and classes that obey restrictions or e.g. are the result of the union of other classes.

2.6 Interviews with specialists and literature consultation

The standard approach to design an ontology, according to the literature, is to interview specialists of the field to which the ontology is related, or to absorb the established literature, or both. This work is being developed using both approaches. The second author is a data visualization specialist which was interviewed by the first author. Also, the first author is acquiring a deeper knowledge in the field.

3. RESULTS AND DISCUSSION

Using the framework exposed in the previous section, we elaborated an initial vocabulary and ontology for visual analytics: AAVO (Audiovisual Analytics Vocabulary and Ontology). The inclusion of "audio" is both a reminder of the possibilities available for using audio to represent data and perceive patterns, and a fitting incorporation of audio to visual analytics given audiovisual capabilites of current ordinary computers.

[1] Visit `http://dbpedia.org/page/Rhesus_macaque` and click on the concepts and properties to start browsing the web of linked data.

The main concepts and their interrelation are presented in Section 3.1 while an example extension is on Section 3.2. Section 3.3 hold annotations for the vocabulary which are not promptly given by the previous sections. Some of the relations bellow were expressed by very recent techniques that are described in Fabbri (2017). Their meaning, though, might be easily inferred.

3.1 AAVO core

The core of AAVO is designed to be minimal and hold the following concepts as depicted in Figure 1:

- Visualization: a technique to generate a Visual Representation from Data.
- Visual Representation: a representation of data by visual cues. A Visual Representation can be an Image or an Animation.
- Data: a set o values be them qualitative or quantitative (Wikipedia, 2017).
- Dataset Type: a type of organization and meaning of data Munzner (2014).
- Processing: transforms Data into Data. Pre-processing is a kind of Processing.

We envision that there should be at least the following concepts in AAVO core when it reaches maturity:

- Hypothesis: a proposed explanation for a phenomenon that might be 1) given beforehand for being proved or refuted by an Analysis, 2) shaed by means of an Analysis, or 3) presented by means of a Visualization.
- Analysis: a set of procedures used to obtain understanding about Data or a phenomenon.
- Task/Purpose/Application: the goal or objective of an Analysis.

These concepts were not included in AAVO core (e.g. Figure 1) because we chose to better conceptualize the field in order to avoid too much rework. From the definitions above, raises the question: should we also include Phenomenon among these core concepts?

Other relations that can be added to the core or to an extension but are directly related to the core:

- Visualization is a type of Processing.
- Visual Representation is a Dataset Type.
- Processing “suitable for” Data.
- Visualization “number of dimensions” real (not double as stated for now).

An example of question still left unanswered: is a Visualization only outputs Visual Representation or can it output other Data? This and many other questions might have a resolution that are genuinely dependent on the conceptual design of the ontology.

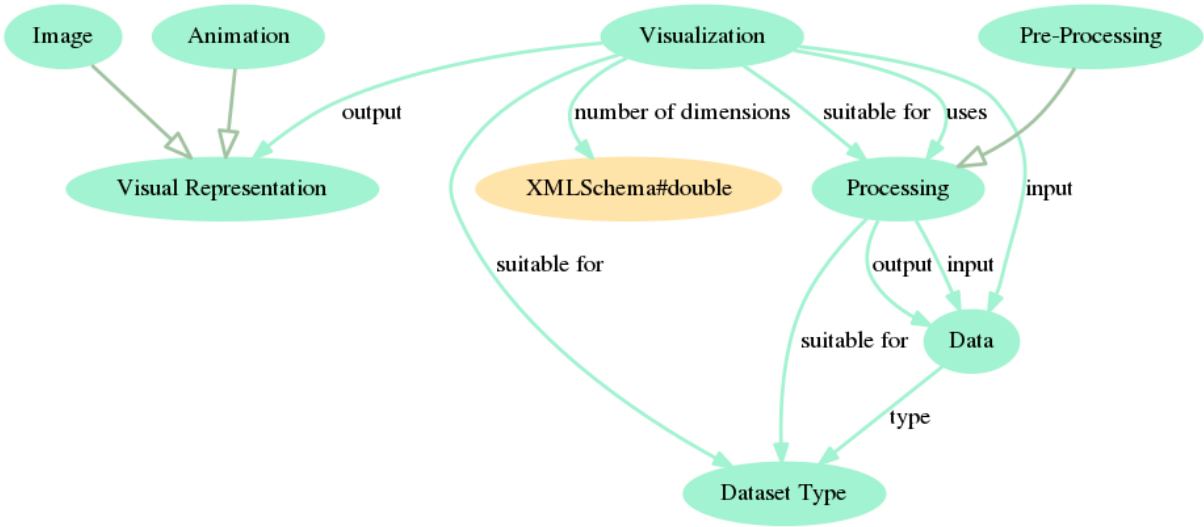

Figure 1: AAVO core, discussed in Section 3.1. A Visualization outputs a Visual Representation that can be an Image or an Animation. A Visualization is suitable for a Dataset Type and for a Processing routine which transforms Data into Data.

3.2 AAVO example expansion

There are many ways in which the AAVO core might be expanded. Figure 2 is an example expansion. Concepts were added which are hyponyms to Dataset Type (Temporal Series, Relational Data), to Pre-Processing (Z-Score, Cleaning), Processing (MDS, Statistical Test) and Visualization (Heat Map, Histogram, Scatter Plot, Timeline). Some example of further subclasses are also added. A different kind of expansion was achieved by including (Data) Availability and the less general concepts of Dynamic Availability and Static Availability. A Graph is regarded as a Network without any context or further attributes beyond nodes and edges.

Ideally, AAVO expansions should reach related fields, such as HCI, by linking to other existing ontologies (such as DBPedia) or incorporating enough concepts to then bind and rely in third party conceptualizations.

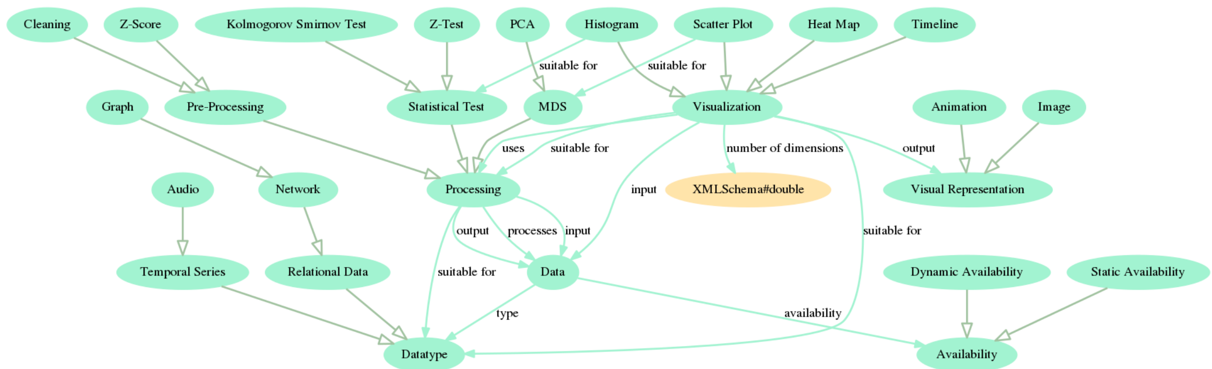

Figure 2: Added less general concepts: Statistical Test, MDS, Timeline, Z-Score, Network, etc. A thorough consideration of this expansion of the AAVO core is at Section 3.2.

3.3 Vocabulary annotations

Beyond what is made explicit in the previous sections, there are some aspects of the knowledge and language that are to be directly added to the SKOS vocabulary. Examples:

- in a dataset, an element is also called: an item, an observation, an individual, a point, and even a data point and a data row.
- A graph node is also called: a vertex, and every name that are used to designate an element.
- A graph edge is also called: a link, a bond, a line, and a connection.
- Z-scores are also called: standard scores, normal scores, standardized variables, and z-values.

4. CONCLUSIONS AND FURTHER WORK

This initial formalized conceptualization of the AAVO holds some relations which are not explicitly described by current literature mainly because: 1) AAVO was created with the purpose of reaching a sound conceptualization that allows it to be formalized as linked data; 2) AAVO has the purpose of representing the knowledge in Visual Analytics to enable inference by machines. There are other uses for AAVO which were uncovered in Section 1, but for those there are conceptual models available Munzner (2014).

Potential further steps include:

- the inclusion of the concepts Hypothesis, Analysis, and Task into the AAVO Core.
- Realizing AAVO expansions until the reached concepts can be linked to other ontologies that are sound, used and maintained.
- Using AAVO for obtaining interesting relations by means of automated inference and for assisting a (audio) visual analytics software.

Acknowledgements

The authors thank CNPq and FAPESP for the funding received while researching the topic of this article, the researchers of IFSC/USP and ICMC/USP for the recurrent collaboration in every situation where we needed directions for investigation.

REFERENCES

Fabbri, R. (2017). Necessary enhancements of linked data expressiveness for ontologies. Encontro Nacional de Modelagem Computacional 2017 (XX ENMC). From `https://github.com/ttm/ontologyEnhancements/raw/master/article.pdf`

Heath, T. & Bizer, C. (2011). Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.

Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., ... & Bizer, C. (2015). DBpedia–a large-scale, multilingual knowledge base extracted from Wikipedia. Semantic Web, 6(2), 167-195.

Munzner, T. (2014). Visualization analysis and design. CRC press.

W3C (2017). LINKED DATA CURRENT STATUS, from https://www.w3.org/standards/techs/linkeddata

Data. (2017, August 21). In Wikipedia, The Free Encyclopedia. Retrieved 22:31, August 21, 2017 , from https://en.wikipedia.org/w/index.php?title=Data&oldid=796493851